# Project Overview: Pet Insurance Analytics Portfolio

## Project Summary

This project simulates the operations of a UK-based pet insurance company using a fully relational MySQL database. It models realistic customer, pet, policy, and claims data, generated synthetically using Python. The goal is to demonstrate the full analytics workflow — from data modeling and SQL analysis to business intelligence and insight storytelling — in a format that mirrors real-world analyst responsibilities.

## Objectives

- Design a normalized relational schema
- Generate realistic synthetic data
- Load and validate data in MySQL
- Write analytical SQL queries
- Create processed, analysis-ready datasets
- Explore data in Python
- Build dashboards in Power BI or Tableau
- Present insights clearly for recruiters and stakeholders

## Database Schema

The schema includes 8 core tables:

- customers: Customer demographics and contact info
- pets: Pet profiles linked to customers
- products: Insurance products and pricing
- policies: Policy subscriptions per pet
- claims: Claims submitted under policies
- vet_visits: Vet appointments linked to claims
- invoice_line_items: Itemized treatment costs

- claim_payments: Payments issued for approved claims

## Key Business Questions

- Which pet breeds generate the highest average claim costs?
- How does pet age affect policy risk and claim frequency?
- What is the loss ratio by product and coverage type?
- What is the average customer lifetime value (CLV)?
- Which vet clinics are most expensive or most visited?
- What is the claim approval rate by diagnosis?
- How many customers own multiple pets or policies?
- What is the average time from claim to payment?

## Processed Datasets

To support analysis and dashboarding, the following views were created:

- fact_claims_summary: Flattened claims + policy + pet + customer
- dim_customers: Customer-level metrics and totals
- dim_products: Product-level performance and loss ratios
- fact_vet_costs: Line-level vet treatment costs

These were exported to CSV using a Python script and stored in /data/processed/.

## Tools & Technologies

- **MySQL** – schema design, data validation, SQL analysis
- **Python** – data generation, export automation
- **Pandas / Seaborn** – EDA and visualisation
- **Power BI / Tableau** – dashboards and reporting
- **GitHub** – version control and documentation